

WHAT IS CLAIMED IS:

1. A system for optimizing speech recognition procedures, comprising:
initial language models each created by combining source models
5 according to interpolation coefficients that define proportional
relationships for combining said source models;
a speech recognizer that utilizes said initial language models to process
input development data for calculating word-error rates that each
correspond to a different one of said initial language models; and
10 an optimized language model selected from said initial language models
by identifying an optimal word-error rate from among said word-
error rates, said speech recognizer utilizing said optimized
language model for performing said speech recognition
procedures.
15
2. The system of claim 1 wherein said word-error rates are calculated by
comparing a correct transcription of said input development data and a top
recognition candidate from an N-best list that is rescored by a rescoring
module for each of said initial language models.
20
3. The system of claim 1 wherein said initial language models are
implemented as statistical language models that include N-grams and
probability values that each correspond to one of said N-grams.
- 25 4. The system of claim 1 wherein said input development data includes
a pre-defined series of word sequences from which said recognizer rescores a
corresponding N-best list for calculating said word-error rates.
5. The system of claim 1 wherein said source models are each similarly
30 implemented as statistical language models that include N-grams and
probability values that each correspond to one of said N-grams.

6. The system of claim 1 wherein each of said source models corresponds to a different application domain that is related to a particular speech environment.

5 7. The system of claim 1 wherein sets of said interpolation coefficients are each associated with a different one of said source models to define how much said different one of said source models contributes to a corresponding one of said initial language models.

10 8. The system of claim 1 wherein said interpolation coefficients are each multiplied with a different one of said source models to produce a series of weighted source models that are then combined to produce a corresponding one of said initial language models.

15 9. The system of claim 1 wherein said initial language models are each calculated by a formula:

$$LM = \lambda_1 SM_1 + \lambda_2 SM_2 + \dots + \lambda_n SM_n$$

20 where said LM is one of said initial language models, said SM_1 is a first one of said source models, said SM_n is a final one of said source models in a continuous sequence of “n” source models, and said λ_1 , said λ_2 , and said λ_n are said interpolation coefficients applied to respective probability values of said source models to weight how much each of said source models
25 contributes to said one of said initial language models.

10. The system of claim 1 wherein said interpolation coefficients are each greater than or equal to “0”, and are also each less than or equal to “1”, a sum of all of said interpolation coefficients being equal to “1”.

30

11. The system of claim 1 wherein said interpolation coefficients for creating said optimized language model are selectively chosen by analyzing effects of various combinations of said interpolation coefficients upon said word-error rates that correspond to recognition accuracy characteristics of said speech recognizer, said optimized language model being directly implemented by minimizing said optimal word-error rate through a selection of said interpolation coefficients.

12. The system of claim 1 wherein a rescoring module repeatedly processes said input development data to rescore an N-best list of recognition candidates for calculating said word-error rates by comparing a top recognition candidate to said input development data, said recognition candidates each including a recognition result in a text format, and a corresponding recognition score.

13. The system of claim 1 wherein each of said word-error rates are calculated by comparing a correct transcription of said input development data and a top recognition candidate from an N-best list of recognition candidates provided by said speech recognizer after processing said input development data, said top recognition candidate corresponding to a best recognition score from said speech recognizer.

14. The system of claim 1 wherein said word-error rates are calculated to include one or more substitutions in which a first incorrect word has been substituted for a first correct word in a recognition result, said word-error rates also including one or more deletions in which a second correct word has been deleted from said recognition result, said word-error rates further including one or more insertions in which a second incorrect word has been inserted into said recognition result.

15. The system of claim 1 wherein said word-error rates are each calculated according to a formula:

$$\text{WER} = (\text{Subs} + \text{Deletes} + \text{Inserts}) / \text{Total Words in Correct Transcription}$$

where said WER is one of said word-error rates corresponding to one of said initial language models, said Subs are substitutions in a recognition result, said Deletes are deletions in said recognition result, said Inserts are insertions in said recognition result, and said Total Words in Correct Transcription is a total number of words in a correct transcription of said input development data.

16. The system of claim 1 wherein an interpolation procedure for combining said source models into one of said initial language models is performed by utilizing a selected initial set of said interpolation coefficients.

17. The system of claim 16 wherein a rescoring module rescores an N-best list of recognition candidates after utilizing said one of said initial language models to perform a recognition procedure upon said input development data.

18. The system of claim 17 wherein one of said word-error rates corresponding to said one of said initial language models is calculated and stored based upon a comparison between a correct transcription of said input development data and a top recognition candidate from said N-best list.

19. The system of claim 18 wherein said selected initial set of said interpolation coefficients are each iteratively altered by a pre-defined amount to produce subsequent sets of said interpolation coefficients.

20. The system of claim 19 wherein subsequent initial language models are created by utilizing said subsequent sets of interpolation coefficients, a rescoring module iteratively utilizing said subsequent initial language models to rescore said N-best list for calculating subsequent word-error rates, said
5 optimized language model being selected by identifying said optimal word-error rate when a pre-determined number of said subsequent word-error rates have been calculated.

21. A method for optimizing speech recognition procedures, comprising:
10 creating initial language models by combining source models according to interpolation coefficients that define proportional relationships for combining said source models;
utilizing said initial language models to process input development data for calculating word-error rates that each correspond to a
15 different one of said initial language models;
selecting an optimized language model from said initial language models by identifying an optimal word-error rate from among said word-error rates; and
utilizing said optimized language model for performing said speech
20 recognition procedures with a speech recognizer.

22. The method of claim 21 wherein said word-error rates are calculated by comparing a correct transcription of said input development data and a top recognition candidate from an N-best list that is rescored by a rescoring
25 module for each of said initial language models.

23. The method of claim 21 wherein said initial language models are implemented as statistical language models that include N-grams and probability values that each correspond to one of said N-grams.
30

24. The method of claim 21 wherein said input development data includes a pre-defined series of word sequences from which said recognizer rescores a corresponding N-best list for calculating said word-error rates.

5 25. The method of claim 21 wherein said source models are each similarly implemented as statistical language models that include N-grams and probability values that each correspond to one of said N-grams.

26. The method of claim 21 wherein each of said source models
10 corresponds to a different application domain that is related to a particular speech environment.

27. The method of claim 21 wherein sets of said interpolation coefficients are each associated with a different one of said source models to define how
15 much said different one of said source models contributes to a corresponding one of said initial language models.

28. The method of claim 21 wherein said interpolation coefficients are each multiplied with a different one of said source models to produce a series of
20 weighted source models that are then combined to produce a corresponding one of said initial language models.

25

30

29. The method of claim 21 wherein said initial language models are each calculated by a formula:

$$LM = \lambda_1 SM_1 + \lambda_2 SM_2 + \dots + \lambda_n SM_n$$

5

where said LM is one of said initial language models, said SM_1 is a first one of said source models, said SM_n is a final one of said source models in a continuous sequence of “n” source models, and said λ_1 , said λ_2 , and said λ_n are said interpolation coefficients applied to respective probability values of said source models to weight how much each of said source models contributes to said one of said initial language models.

30. The method of claim 21 wherein said interpolation coefficients are each greater than or equal to “0”, and are also each less than or equal to “1”, a sum of all of said interpolation coefficients being equal to “1”.

31. The method of claim 21 wherein said interpolation coefficients for creating said optimized language model are selectively chosen by analyzing effects of various combinations of said interpolation coefficients upon said word-error rates that correspond to recognition accuracy characteristics of said speech recognizer, said optimized language model being directly implemented by minimizing said optimal word-error rate through a selection of said interpolation coefficients.

32. The method of claim 21 wherein a rescoring module repeatedly processes said input development data to generate and rescore an N-best list of recognition candidates for calculating said word-error rates by comparing a top recognition candidate to said input development data, said recognition candidates each including a recognition result in a text format, and a corresponding recognition score.

33. The method of claim 21 wherein each of said word-error rates are calculated by comparing a correct transcription of said input development data and a top recognition candidate from an N-best list of recognition candidates provided by said speech recognizer after processing said input development data, said top recognition candidate corresponding to a best recognition score from said speech recognizer.

34. The method of claim 21 wherein said word-error rates are calculated to include one or more substitutions in which a first incorrect word has been substituted for a first correct word in a recognition result, said word-error rates also including one or more deletions in which a second correct word has been deleted from said recognition result, said word-error rates further including one or more insertions in which a second incorrect word has been inserted into said recognition result.

35. The method of claim 21 wherein said word-error rates are each calculated according to a formula:

$$\text{WER} = (\text{Subs} + \text{Deletes} + \text{Inserts}) / \text{Total Words in Correct Transcription}$$

where said WER is one of said word-error rates corresponding to one of said initial language models, said Subs are substitutions in a recognition result, said Deletes are deletions in said recognition result, said Inserts are insertions in said recognition result, and said Total Words in Correct Transcription is a total number of words in a correct transcription of said input development data.

36. The method of claim 21 wherein an interpolation procedure for combining said source models into one of said initial language models is performed by utilizing a selected initial set of said interpolation coefficients.

37. The method of claim 36 wherein a rescoring module rescores an N-best list of recognition candidates after utilizing said one of said initial language models to perform a recognition procedure upon said input development data.

5

38. The method of claim 37 wherein one of said word-error rates corresponding to said one of said initial language models is calculated and stored based upon a comparison between a correct transcription of said input development data and a top recognition candidate from said N-best list.

10

39. The method of claim 38 wherein said selected initial set of said interpolation coefficients are each iteratively altered by a pre-defined amount to produce subsequent sets of said interpolation coefficients.

15 40. The method of claim 39 wherein subsequent initial language models are created by utilizing said subsequent sets of interpolation coefficients, a rescoring module iteratively utilizing said subsequent initial language models to rescore said N-best list for calculating subsequent word-error rates, said optimized language model being selected by identifying said optimal word-error rate when a pre-determined number of said subsequent word-error rates have been calculated.

20

25

30

41. A system for optimizing speech recognition procedures, comprising:
means for creating initial language models by combining source models
according to interpolation coefficients that define proportional
relationships for combining said source models;
5 means for utilizing said initial language models to process input
development data for calculating word-error rates that each
correspond to a different one of said initial language models;
means for selecting an optimized language model from said initial
language models by identifying an optimal word-error rate from
10 among said word-error rates; and
means for utilizing said optimized language model for performing said
speech recognition procedures.

42. A system for optimizing speech recognition procedures, comprising:
15 initial language models each created by combining source models
according to interpolation coefficients that define proportional
relationships for combining said source models;
a speech recognizer that utilizes said initial language models to process
input development data for calculating word-error rates that each
20 correspond to a different one of said initial language models, said
word-error rates being calculated by comparing a correct
transcription of said input development data and a top
recognition candidate from an N-best list that is rescored by a
rescoring module for each of said initial language models; and
25 an optimized language model selected from said initial language models
by identifying an optimal word-error rate from among said word-
error rates, said speech recognizer utilizing said optimized
language model for performing said speech recognition
procedures.